Crawling and Gathering URLs(Gospider, Katana , Waybackurls)

- تجميع الروابط الخاصه بالمواقع وازاي تصنف الملفات وتحط كل حاجه لوحدها

```
cat subdomains.txt => subfinder , assetfinder - security trials , subdomain
finder c99 .nl , eyeshrew

cat subdomains.txt | httpx -o httpx.txt


-> gather url => https://sub.example.com =>
https://sub.example.com/file.php?url=x&id=1

Passive =>
Active =>

        - Waybackurls
wayback machine => internet archive
        - Katana => 2022 => # katana -list httpx -o allurls.txt20
        - Gospider
--------------------------
```

- اول حاجه انت بتجمع ال subdomains من المواقع والادوات
- بعدا ما تحطهم ف ملف بتستخدم ال httpx علشان تشوف الروابط اللي شغاله
- بتجمع روابط ودا بيبقي شكل الرابط
- https://sub.example.com/file.php?url=x&id=1
- في نوعين من التجميع
  - الـ Passive
    - بتجمع معلومات عن الموقع من مواقع مختلفه خارجيه من غير ما تحتك بالموقع
  - الـ Active
    - بتحتك بالموقع بشكل مباشر زي ال fuzzing اللي بتعملو ع الموقع

```
        - Waybackurls
wayback machine => internet archive
        - Katana => 2022 => # katana -list httpx -o allurls.txt20
        - Gospider
--------------------------

cat urls1.txt | grep -E "\.js" >> js.txt
cat urls1.txt | grep -E "\.php" >> php.txt
```

- اول اداة اسمها waybackurls
  - بتنزلها وبتلاقيها ف مسار go/bin/
- انت بتبقي حاطط ال subdomain ف فايل ف بتستخدمه
- cat subs.txt | waybackurls >> wayback.txt
- وبعدها اسم الملف اللي هتحفظهم فيه
- بتجمعلك الروابط من موقع internet archive

- الاداة التانيه اسمها katana
- بتشتغل كدا
- katana -list subs.txt -o allutls.txt
  - بعد -list بتحط الملف
  - بعد - o بتحط اسم الملف اللي هتحفظهم فيه

- التالته اسمها Gospider
  - لو هتشتغل علي الاداة ب link و احد بس بتحطلها
- -s
  - لكن لو معاك ملف فيه list مواقع بتستخدم
- -S
- طريقه استخدامها
- gospider -S sub.txt -o gospider
  - بتحفظها ف Folder مش file

```
[url] - [code-200] - https://0120a109484c0b1abae707a05aa3693b.banfield.us
[href] - https://banfieldmaintenance.blob.core.windows.net/images/favicon.ico
[href] - https://stackpath.bootstrapcdn.com/bootstrap/4.4.1/css/bootstrap.min.css
[href] - https://stackpath.bootstrapcdn.com/font-awesome/4.7.0/css/font-awesome.min.css
[href] - https://cdnjs.cloudflare.com/ajax/libs/bootstrap-datepicker/1.9.0/css/bootstrap-
[href] - https://0120a109484c0b1abae707a05aa3693b.banfield.us
[url] - [code-200] - https://61e34e85801b71354676ae1cce02da91.banfield.us
[href] - https://banfieldmaintenance.blob.core.windows.net/images/favicon.ico
[href] - https://stackpath.bootstrapcdn.com/bootstrap/4.4.1/css/bootstrap.min.css
[href] - https://stackpath.bootstrapcdn.com/font-awesome/4.7.0/css/font-awesome.min.css
[href] - https://cdnjs.cloudflare.com/ajax/libs/bootstrap-datepicker/1.9.0/css/bootstrap-
[href] - https://61e34e85801b71354676ae1cce02da91.banfield.us
[url] - [code-200] - https://620239893ed9a60001143aca.banfield.us
[href] - https://banfieldmaintenance.blob.core.windows.net/images/favicon.ico
```

- الرابط بيبقي بعده وقبله كلام علشان تحذفه بتعمل

- 1- ctrl+w
- 2- ctrl+r Replace
- 3- alt+r Reg.exp

- علشان تحذفهم بتكتب هتحذف ال [url]
  - بتحط ف الاول / بعدها اللي عايز تحذفه

- \[url\]... enter enter a
  - النقط دي بتبقي علي حسب اللي بعدها المسافه بنقطه لو في - بتتحسب نقطه
- بتعيد كل الخطوات لحد ما كلو يتمسح

دا كدا الكلام اللي ف الاول

- اللي ف الاخر

- ctrl+w
- ctrl+r

- بتحدد من القوس للاخر

- \].+ enter enter a

- بعد دا كلو بتقرا ال3 ملفات بتوع ال 3 ادوات وتحطهم ف فايل واحد

- cat wayback.txt katana.txt gospinder.txt | anew >> urls.txt
  - anew علشان تحذف المقرر

```
cat urls1.txt | grep -E "\.js" >> js.txt
cat urls1.txt | grep -E "\.php" >> php.txt
```

- اخر حاجه بتحط ال js ف فايل وال php ف فايل